Fisher Auto-Encoders: Supplementary Materials

1 Proof of Theorem 1

$$
\begin{aligned}
\phi^\star, \eta^\star, \theta^\star &= \arg\min_{\phi,\eta,\theta} \mathbb{D}_\nabla \left[q_{\star,\phi}(\mathbf{x},\mathbf{z}) || p_\theta(\mathbf{x},\mathbf{z}) \right] \\
&= \arg\min_{\phi,\eta,\theta} \mathbb{E}_{q_{\star,\phi}(\mathbf{x},\mathbf{z})} \frac{1}{2} \left\| \nabla_{\mathbf{x},\mathbf{z}} \log q_{\star,\phi}(\mathbf{x},\mathbf{z}) - \nabla_{\mathbf{x},\mathbf{z}} \log p_{\eta,\theta}(\mathbf{x},\mathbf{z}) \right\|^2 \\
&= \arg\min_{\phi,\eta,\theta} \mathbb{E}_{p_\star(\mathbf{x})} \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} \frac{1}{2} \left\| \nabla_{\mathbf{x},\mathbf{z}} \log q_{\star,\phi}(\mathbf{x},\mathbf{z}) - \nabla_{\mathbf{x},\mathbf{z}} \log p_{\eta,\theta}(\mathbf{x},\mathbf{z}) \right\|^2 \\
&= \arg\min_{\phi,\eta,\theta} \mathbb{E}_{p_\star(\mathbf{x})} \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} \frac{1}{2} \left\| \nabla_{\mathbf{x}} \log p_\star(\mathbf{x}) + \nabla_{\mathbf{x}} \log q_\phi(\mathbf{z}|\mathbf{x}) - \nabla_{\mathbf{x}} \log p_\theta(\mathbf{x}|\mathbf{z}) \right\|^2 \\
&\qquad + \mathbb{E}_{p_\star(\mathbf{x})} \underbrace{\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} \frac{1}{2} \left\| \nabla_{\mathbf{z}} \log q_\phi(\mathbf{z}|\mathbf{x}) - \nabla_{\mathbf{z}} \log p_{\eta,\theta}(\mathbf{z}|\mathbf{x}) \right\|^2}_{\mathbb{D}_\nabla [q_\phi(\mathbf{z}|\mathbf{x}) || p_{\eta,\theta}(\mathbf{z}|\mathbf{x})]} \\
&= \arg\min_{\phi,\eta,\theta} \mathbb{E}_{p_\star(\mathbf{x})} \mathbb{D}_\nabla \left[q_\phi(\mathbf{z}|\mathbf{x}) || p_{\eta,\theta}(\mathbf{z}|\mathbf{x}) \right] + \mathbb{E}_{p_\star(\mathbf{x})} \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} \frac{1}{2} \left\| \nabla_{\mathbf{x}} \log p_\star(\mathbf{x}) \right\|^2 \\
&\qquad + \mathbb{E}_{p_\star(\mathbf{x})} \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} \nabla_{\mathbf{x}} \log p_\star(\mathbf{x})^\top \nabla_{\mathbf{x}} \log q_\phi(\mathbf{z}|\mathbf{x}) \\
&\qquad + \mathbb{E}_{p_\star(\mathbf{x})} \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} \frac{1}{2} \left\| \nabla_{\mathbf{x}} \log q_\phi(\mathbf{z}|\mathbf{x}) \right\|^2 - \nabla_{\mathbf{x}} \log p_\theta(\mathbf{x}|\mathbf{z})^\top \nabla_{\mathbf{x}} \log q_\phi(\mathbf{z}|\mathbf{x}) \\
&\qquad + \mathbb{E}_{p_\star(\mathbf{x})} \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} \frac{1}{2} \left\| \nabla_{\mathbf{x}} \log p_\theta(\mathbf{x}|\mathbf{z}) \right\|^2 - \nabla_{\mathbf{x}} \log p_\theta(\mathbf{x}|\mathbf{z})^\top \nabla_{\mathbf{x}} \log p_\star(\mathbf{x})
\end{aligned}
$$

Let's examine the inner-product terms:

$$
\begin{aligned}
&\mathbb{E}_{p_\star(\mathbf{x})} \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} \nabla_{\mathbf{x}} \log p_\star(\mathbf{x})^\top \nabla_{\mathbf{x}} \log q_\phi(\mathbf{z}|\mathbf{x}) \\
&= \iint p_\star(\mathbf{x}) q_\phi(\mathbf{z}|\mathbf{x}) \nabla_{\mathbf{x}} \log p_\star(\mathbf{x})^\top \nabla_{\mathbf{x}} \log q_\phi(\mathbf{z}|\mathbf{x}) d\mathbf{z} d\mathbf{x} \\
&= \iint p_\star(\mathbf{x}) q_\phi(\mathbf{z}|\mathbf{x}) \nabla_{\mathbf{x}} \log p_\star(\mathbf{x})^\top \nabla_{\mathbf{x}} \log q_\phi(\mathbf{z}|\mathbf{x}) d\mathbf{x} d\mathbf{z} \\
&= \iint p_\star(\mathbf{x}) \nabla_{\mathbf{x}} \log p_\star(\mathbf{x})^\top \nabla_{\mathbf{x}} q_\phi(\mathbf{z}|\mathbf{x}) d\mathbf{x} d\mathbf{z} \\
&\overset{(a)}{=} - \iint p_\star(\mathbf{x}) q_\phi(\mathbf{z}|\mathbf{x}) \left[\left\| \nabla_{\mathbf{x}} \log p_\star(\mathbf{x}) \right\|^2 + \Delta_{\mathbf{x}} \log p_\star(\mathbf{x}) \right] d\mathbf{x} d\mathbf{z} \\
&= -\mathbb{E}_{p_\star(\mathbf{x})} \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} \left[\left\| \nabla_{\mathbf{x}} \log p_\star(\mathbf{x}) \right\|^2 + \Delta_{\mathbf{x}} \log p_\star(\mathbf{x}) \right],
\end{aligned}
$$

where (a) is obtained by an integration by parts.

$$\begin{aligned}
&\mathbb{E}_{p_\star(\mathbf{x})}\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} - \nabla_{\mathbf{x}} \log p_\theta(\mathbf{x}|\mathbf{z})^\top \nabla_{\mathbf{x}} \log q_\phi(\mathbf{z}|\mathbf{x}) \\
&= -\iint p_\star(\mathbf{x}) q_\phi(\mathbf{z}|\mathbf{x}) \nabla_{\mathbf{x}} \log p_\theta(\mathbf{x}|\mathbf{z})^\top \nabla_{\mathbf{x}} \log q_\phi(\mathbf{z}|\mathbf{x}) d\mathbf{z} d\mathbf{x} \\
&= -\iint p_\star(\mathbf{x}) \nabla_{\mathbf{x}} \log p_\theta(\mathbf{x}|\mathbf{z})^\top \nabla_{\mathbf{x}} q_\phi(\mathbf{z}|\mathbf{x}) d\mathbf{x} d\mathbf{z} \\
&\overset{(b)}{=} \iint p_\star(\mathbf{x}) q_\phi(\mathbf{z}|\mathbf{x}) \left[\Delta_{\mathbf{x}} \log p_\theta(\mathbf{x}|\mathbf{z}) + \nabla_{\mathbf{x}} \log p_\theta(\mathbf{x}|\mathbf{z})^\top \nabla_{\mathbf{x}} \log p_\star(\mathbf{x})\right] d\mathbf{x} d\mathbf{z} \\
&= \mathbb{E}_{p_\star(\mathbf{x})}\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} \left[\Delta_{\mathbf{x}} \log p_\theta(\mathbf{x}|\mathbf{z}) + \nabla_{\mathbf{x}} \log p_\theta(\mathbf{x}|\mathbf{z})^\top \nabla_{\mathbf{x}} \log p_\star(\mathbf{x})\right],
\end{aligned}$$

where (b) is again obtained by an integration by parts. Grouping all the terms together, we get

$$\begin{aligned}
&\phi^\star, \eta^\star, \theta^\star \\
&= \arg\min_{\phi,\eta,\theta} \mathbb{E}_{p_\star(\mathbf{x})} - s_\nabla \left[p_\star(\mathbf{x})\right] + \mathbb{D}_\nabla \left[q_\phi(\mathbf{z}|\mathbf{x}) || p_{\eta,\theta}(\mathbf{z}|\mathbf{x})\right] + \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} s_\nabla \left[p_\theta(\mathbf{x}|\mathbf{z})\right] \\
&\qquad + \frac{1}{2} \left\|\nabla_{\mathbf{x}} \log q_\phi(\mathbf{z}|\mathbf{x})\right\|^2.
\end{aligned}$$

By noticing that $\mathbb{E}_{p_\star(\mathbf{x})} - s_\nabla \left[p_\star(\mathbf{x})\right]$ is independent of the parameters ϕ, η and θ, we conclude the proof of Theorem 1.

2 Robustness to binary masking noise

We extend the experiments to examine the robustness of the proposed Fisher AEs and consider another type of noise called binary masking noise which consists on setting the value of a randomly selected fraction ν of input components to zero.

2.1 MNIST

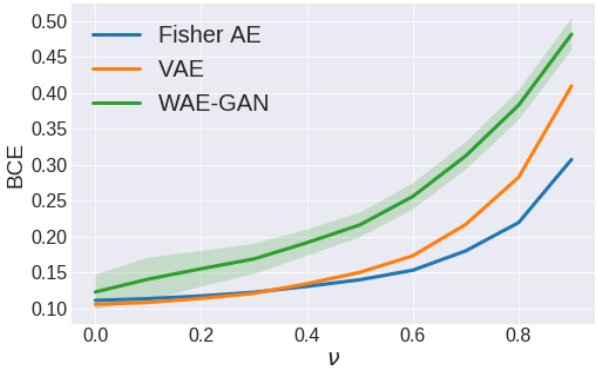

Figure 1: BCE vs. the fraction ν for MNIST.

The same insights regarding the robustness of the Fisher AE to Gaussian noise are confirmed in the case of binary masking noise. Both Figures 1 and 2 shows the superiority of Fisher AE in terms of robustness to binary masking noise as compared to VAE and WAE-GAN.

2.2 CelebA

As shown in Figures 3 and 4, in the case of CelebA, both VAE and Fisher AE exhibit similar but superior performance in terms of robustness against binary masking noise as compared to WAE-GAN.

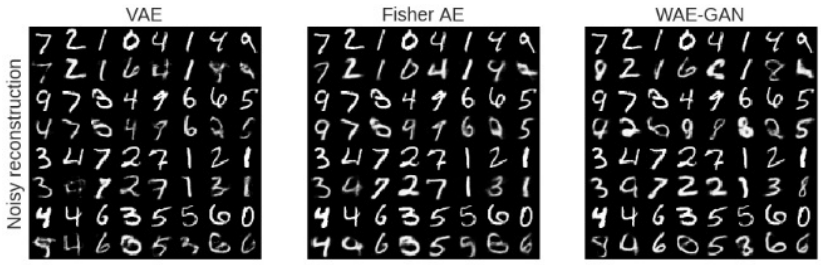

Figure 2: Test reconstruction results when a random fraction $\nu = 0.8$ of test data is set to zero. True test data are given by the odd rows.

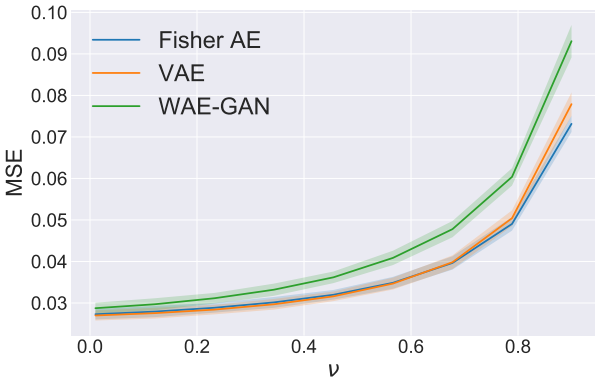

Figure 3: MSE vs. the fraction ν for celebA with random mask.

3 Further details on experiments

Here, we give the detailed architecture used in the implementation of the different auto-encoders for both MNIST and celebA data sets.

- FC(n_{in}, n_{out}): Fully connected layer with input/output dimensions given by n_{in} and n_{out}.
- Conv(n_{in}, n_{out}, k, s, p): Convolutional layer with input channels n_{in}, output channels n_{out}, kernel size k, stride s and padding p.
- ConvT(n_{in}, n_{out}, k, s, p): Transposed convolutional layer with input channels n_{in}, output channels n_{out}, kernel size k, stride s and padding p.
- AvgPool(k, s, p): Average Pooling with kernel size, stride and padding respectively given by k, s and p.
- BN : Batch-normalization
- BiI : 2D bilinear interpolation layer

Figure 4: Test reconstruction results when a random fraction $\nu = 0.8$ of test data is set to zero. True test data are given by the odd rows.

3.1 MNIST

Encoder	**Decoder**
Input size: (1, 28, 28) Conv(1, 64, 3, 2, 2) LeakyReLU Conv(64, 128, 3, 2, 2) BN LeakyReLU Conv(128, 256, 3, 2, 2) BN, LeakyReLU Conv(256, 512, 3, 2, 1) BN, LeakyReLU Conv(512, 16, 3, 2, 0) Output size : (16, 1, 1)	Input size: (8, 1, 1) ConvT(8, 512, 5, 1, 1) BN, ReLU ConvT(512, 256, 5, 1, 1) BN, ReLU ConvT(256, 128, 5, 2, 1) BN, ReLU ConvT(128, 64, 5, 1, 1) BN, ReLU ConvT(64, 1, 4, 2, 0) Sigmoid Output size : (1, 28, 28)

Table 1: Encoder/Decoder architectures for Fisher AE and VAE for MNIST

Encoder	**Generator**	**Discriminator**
Input size : (1, 28, 28) Conv(1, 64, 3, 2, 2) LeakyReLU Conv(64, 128, 3, 2, 2) BN, LeakyReLU Conv(128, 256, 3, 2, 2) BN, LeakyReLU Conv(256, 512, 3, 2, 1) BN, LeakyReLU Conv(512, 8, 3, 1, 0) Output size : (8, 1, 1)	Input size : (8, 1, 1) ConvT(8, 512, 5, 1, 1) BN, ReLU ConvT(512, 256, 5, 1, 1) BN, ReLU ConvT(256, 128, 5, 2, 1) BN, ReLU ConvT(128, 64, 5, 1, 1) BN, ReLU ConvT(64, 1, 4, 2, 0) Sigmoid Output size : (1, 28, 28)	Input size : (8, 1, 1) Flatten FC(8, 256) ReLU FC(256, 1) Sigmoid Output size : (1,)

Table 2: Encoder/Generator/Discriminator architectures for WAE-GAN for MNIST

3.2 CelebA

Encoder	**Decoder**
Input size: (3, 64, 64)	Input size: (64, 1, 1)
Conv(3, 64, 5, 1, 2)	BiI, Conv(64, 512, 5, 1, 02)
LeakyReLU, AvgPool(2, 2, 0)	BN, ReLU
Conv(64, 128, 5, 1, 2)	BiI, Conv(512, 256, 5, 1, 2)
BN, LeakyReLU, AvgPool(2, 2, 0)	BN, ReLU
Conv(128, 256, 5, 1, 2)	BiI, Conv(256, 128, 5, 1, 2)
BN, LeakyReLU, AvgPool(2, 2, 0)	BN, ReLU
Conv(256, 512, 5, 1, 2)	BiI, Conv(128, 64, 5, 1, 2)
BN, LeakyReLU	BN, ReLU
Flatten	BiI, Conv(64, 3, 5, 1, 2)
FC(8192, 64)	Tanh
Output size : (64,)	Output size : (3, 64, 64)

Table 3: Encoder/Decoder architectures for Fisher AE and VAE for CelebA

Encoder	**Generator**	**Discriminator**
Input size: (3, 64, 64)	Input size: (64, 1, 1)	
Conv(3, 64, 5, 1, 2)	BiI, Conv(64, 512, 5, 1, 02)	Input size : (8, 1, 1)
LeakyReLU, AvgPool(2, 2, 0)	BN, ReLU	Flatten
Conv(64, 128, 5, 1, 2)	BiI, Conv(512, 256, 5, 1, 2)	FC(8, 256)
BN, LeakyReLU, AvgPool(2, 2, 0)	BN, ReLU	ReLU
Conv(128, 256, 5, 1, 2)	BiI, Conv(256, 128, 5, 1, 2)	FC(256, 1)
BN, LeakyReLU, AvgPool(2, 2, 0)	BN, ReLU	Sigmoid
Conv(256, 512, 5, 1, 2)	BiI, Conv(128, 64, 5, 1, 2)	
BN, LeakyReLU	BN, ReLU	
Flatten	BiI, Conv(64, 3, 5, 1, 2)	Output size : (1,)
FC(8192, 64)	Tanh	
Output size : (64,)	Output size : (3, 64, 64)	

Table 4: Encoder/Generator/Discriminator architectures for WAE-GAN for CelebA